Identification of Protein Abundance Changes in Hepatocellular Carcinoma Tissues Using PCT-SWATH

Journal Article

Author(s):

Zhu, Yi; Zhu, Jiang; Lu, Cong; Zhang, Qiushi; Xie, Wei; Sun, Ping; Dong, Xiaochuan; Yue, Liang; Sun, Yaoting; Yi, Xiao; Zhu, Tiansheng; Ruan, Guan; Aebersold, Rudolf; Huang, Shi'ang; Guo, Tiannan

Publication date:

2019

Permanent link:

<https://doi.org/10.3929/ethz-b-000315521>

Rights / license:

In Copyright - Non-Commercial Use Permitted

Originally published in:

Proteomics. Clinical Applications, <https://doi.org/10.1002/prca.201700179>

Funding acknowledgement:

166435 - MitoModules: Biomarkers in context (SNF)

670821 - Proteomics 4D: The proteome in context (EC)

668858 - PERSONALIZED ENGINE FOR CANCER INTEGRATIVE STUDY AND EVALUATION, a tool for cancer patient risk-stratification and pers. drug selection through multi-omic data integration. (SBFI)

Identification of protein abundance changes in hepatocellular carcinoma tissues using PCT-SWATH

Yi Zhu^{1,2*}, Jiang Zhu^{3*}, Cong Lu³, Qiushi Zhang¹, Wei Xie³, Ping Sun⁴, Xiaochuan Dong⁵, Liang Yue¹, Yaoting Sun¹, Xiao Yi¹, Tiansheng Zhu¹, Guan Ruan¹, Ruedi Aebersold^{2,6#}, Shi'ang Huang^{3#}, Tiannan Guo^{1,2#}

¹ Westlake Institute for Advanced Study, Westlake University, Hangzhou, Zhejiang, P. R. China

² Department of Biology, Institute of Molecular Systems Biology, Eidgenössische Technische Hochschule (ETH) Zurich, Zurich, Switzerland

³ Center for Stem Cell Research and Application, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, P. R. China

⁴ Department of Hepatobiliary Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, P. R. China

⁵ Department of Pathology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, P. R. China

⁶ Faculty of Science, University of Zürich, Zürich, Switzerland

*, co-first authors

Correspondence should be addressed to Ruedi Aebersold (aebersold@imsb.biol.ethz.ch), Shi'ang Huang (sa2huang@hotmail.com), and Tiannan Guo (guotiannan@wias.org.cn)

Abstract

Purpose: To rapidly identify protein abundance changes in biopsy-level fresh-frozen hepatocellular carcinoma (HCC)

Experimental design: We applied the pressure-cycling technology (PCT) and optimized sequential window acquisition of all theoretical mass spectra (SWATH-MS) workflow to analyze 38 biopsy-level tissue samples from 19 HCC patients. Each proteome was analyzed with 45 min liquid chromatography (LC) gradient. MCM7 was validated using immunohistochemistry (IHC).

Received: 03 08, 2018; Revised: 10 16, 2018; Accepted: 10 24, 2018

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/prca.201700179](#).

This article is protected by copyright. All rights reserved.

Results: We quantified 11,787 proteotypic peptides from 2,579 SwissProt proteins with high confidence. The coefficient of variation (CV) of peptide yield using PCT was 32.9%, and the R^2 of peptide quantification was 0.9729. 541 proteins showed significant abundance change between the tumor area and its adjacent benign area. From 24 upregulated pathways and 13 suppressed ones, we observed enhanced biomolecule synthesis and suppressed small molecular metabolism in liver tumor tissues. We further analyzed protein changes based on α -fetoprotein expression and HBV infection. Our data altogether highlighted 16 promising tumor marker candidates. The up-regulation of minichromosome maintenance complex component 7 (MCM7) was further observed in multiple HCC tumor tissues by IHC.

Conclusions and clinical relevance: We demonstrated the practicality of rapid proteomic analysis of biopsy-level fresh-frozen HCC tissue samples with PCT-SWATH and identified promising tumor marker candidates including MCM7.

Clinical Relevance

Better protein biomarkers are required for cancers including HCC. However, quantitative proteomic analysis of tumor tissue samples is of low-throughput.

This study demonstrates the practicality of performing rapid and reproducible proteome analysis of PCT-processed biopsy-level fresh-frozen tumor tissue samples using a single-shot 45-min-LC-gradient SWATH analysis, allowing over 20 proteomes be practically quantified in a working day using a MS instrument. MCM7 is identified as one of the up-regulated proteins in HCC.

The work is a proof-of-principle study exhibiting the ability of the PCT-SWATH methodology to perform quantitative proteome analyses for tissue samples in unprecedented throughput.

Introduction

Hepatocellular carcinoma is the fifth frequent malignancy worldwide and ranks as the third leading cause of cancer-related mortality [1]. In China, HCC leads to more than 300,000 deaths every year [2]. In early-stage disease, HCC patients' survival can be significantly improved by treatment including surgical resection and liver transplant. However, HCC mostly exhibits no symptom at early stages. Consequently, less than one third of newly diagnosed HCC patients are eligible for potential curative therapies. Therefore, early diagnosis of HCC is vital

for patient survival, and the identification of biomarkers for early HCC detection is a crucial clinical need [3].

Currently three FDA approved serum biomarkers are recommended to indicate the risk of liver cancer in high risk populations: AFP, AFP-L3, which is a fucosylated isoform of AFP, and des-gamma-carboxy-prothrombin (DCP) [4]. However, these markers are not included in the surveillance guidelines published by the American Association for the Study of Liver Diseases [5] and the European Association for the Study of the Liver [6] because of low sensitivity and specificity. In China, AFP is used jointly with ultrasonography, computed tomography (CT) and pathology examination for early screening of HCC patients, as recommended by the Asian Pacific Association for the Study of the Liver [7]. This is based on the knowledge that the positive rate of AFP in HCC is about 60-80% [8].

Better biomarkers are needed for HCC. The search for HCC biomarkers from clinical specimens using proteomics approaches mainly focused on blood samples in the past [9, 10]. Besides, a few studies investigated biological fluids including urine [11] and tissue interstitial fluid [12]. To our knowledge, dysregulated tissue proteins are presumably the main source of potential blood and bio-fluidic biomarkers, therefore HCC tissue samples bear great potential in the HCC biomarker studies. Several groups have attempted to identify peptide ions by matrix-assisted laser desorption/ionization (MALDI) imaging mass spectrometry (MS) to classify HCC tissues [13, 14], however, these studies did not characterize the respective proteomes. Li, et al studied malignant and matched benign tissues from eleven HCC patients using two-dimensional liquid chromatography (LC) coupled with shotgun proteomics with a gradient of 165min-LC and observed upregulation of SET complex proteins[15]. In another study, from about 573 proteins identified by 2D gel electrophoresis and label-free shotgun proteomic analysis of 7 pairs of HCC tissue samples, Megger, et al nominated 51 differentially expressed proteins [16]. In a more recent study, Naboulsi W. *et al.* identified and quantified 2,736 proteins from 19 pairs of HCC tumor and adjacent benign tissues using the label-free shotgun proteomics over a 98min-LC gradient [17]. These studies based on shotgun proteomics and 2D gel electrophoresis analysis are generally of low throughput.

We have recently developed a methodology integrating PCT and SWATH-MS for the rapid acquisition of proteotypes, defined as the acute quantitative state of a proteome, from biopsy-level tissue samples [18]. PCT technology produces high pressure (up to 45,000 p.s.i.) and lyses tissue samples by oscillating the pressure between low and high values, thereby

allowing tissue lysis and protein digestion to occur in the same PCT-MicroTube. The method can be applied in a semi-automated and standardized manner and achieves relatively high sample throughput [18-20]. SWATH-MS is an emerging data-independent acquisition (DIA) mass spectrometric method that offers a high degree of quantitative accuracy, proteomic coverage, reproducibility of proteome coverage and sample-throughput [21]. Our previous proteomic study of prostate intra-tissue heterogeneity revealed most thus measured proteins from biopsy-level samples are quantitatively robust in prostate tissues [22]. The PCT-SWATH methodology was further demonstrated to be applicable to rapid proteotype acquisition from sub milligrams of tissue samples [20]. In our first study that demonstrated the technology, the PCT-SWATH method was applied to process and to convert 18 biopsy samples from nine patients with renal cell carcinoma into SWATH-MS fragment ion maps [18]. With a two-hour LC gradient, a 32 fixed SWATH windows setting, and the OpenSWATH data analysis tool [23], about 2000 SwissProt proteins were identified with a high degree of reproducibility across all samples [18]. More recently, a more advanced and miniaturized device, *i.e.* PCT-MicroPestle [19] was introduced providing for higher efficiency of protein extraction and peptide generation from biopsy-level tissues.

In this study, we firstly improved the SWATH methodology by adopting a 45-min LC gradient and an acquisition method using 67 variable SWATH windows to increase the sample throughput without compromising the proteomic depth. Then we applied the improved PCT-SWATH workflow to analyze peptide samples from a cohort of 19 HCC patients in technical duplicate with unprecedented reproducibility and throughput. In total, 11,787 proteotypic peptides from 2,579 SwissProt proteins in 38 HCC tissue samples and 76 SWATH-MS runs were quantified. 16 proteins of high documented relevance to HCC are highlighted in our data with significant regulations in tumor tissues.

Materials and Methods

Patients and tissue samples for SWATH-MS

Tissue specimens from HCC patients were collected from Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. All tissue samples were collected from hepatectomy specimens within one hour after surgical removal, snap frozen and stored at -80°C until proteomic analysis. 38 fresh frozen tissues from 19 HCC patients were included for PCT-SWATH analysis. For each patient, two tissue biopsy punches (with dimensions of 5*5*5 mm³) including a tumor tissue and a benign tissue from an adjacent region

as determined by histomorphological examination were collected. The clinical and pathological data for these patients are provided in Supplementary Table 1. The sample collection was approved by institutional review board of the Union Hospital of Wuhan in China. This study was approved by Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology.

PCT-assisted sample preparation

For each sample, on average about 2mg (weight range: 1.6mg-2.5mg) of the frozen tissue was processed based on the protocol described previously [18, 19, 24]. Briefly, tissues were lysed with the PCT-MicroPestle device in 30μL of lysis buffer composed of 8M urea, 0.1M ammonium bicarbonate, and protease inhibitor (Roche, Switzerland) in a barocycler HUB440 (Pressure BioSciences Inc, South Easton, MA, USA). The tissue lysis was performed under a program consisting of 60 oscillating cycles, with each cycle consisting 50s of high pressure at 45,000psi and 10s of ambient pressure at 32°C. Then the extracted proteins were reduced and alkylated by incubation with 10mM tris (2-carboxyethyl) phosphine (TCEP) and 20mM iodoacetamide (IAA) under gentle vortexing at 600 rpm for 30min (ambient pressure), at 25°C in the dark. Afterwards, digestion was first performed with lys-C (Wako Japan; enzyme-to-substrate ratio, 1:40) in the barocycler (20,000psi, 50s high pressure and 10s ambient pressure, 45 cycles), followed by trypsin (Promega, USA; enzyme-to-substrate ratio, 1:50) digestion in the barocycler (20,000psi, 50s high pressure and 10s ambient pressure, 90 cycles). After digestion, the peptides were acidified with trifluoroacetic acid (TFA) to pH 2-3, and cleaned with SEP-PAK C18 (1cc 50mg) cartridges (Waters, Milford, USA). The peptides were then dried under vacuum (CentriVap, Labconco, Kansas City, USA) and stored at -80 °C before MS analysis. Peptides were dissolved in HPLC-grade water containing 0.1% formic acid and 2% acetonitrile. Peptide concentration was measured with a NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, USA) via absorbance 280 nm (1 Ab = 1 mg/ml).

SWATH-MS on 5600 TripleTOF

The peptide samples were spiked with iRT peptides (Biognosis, Zurich, Switzerland) at a 1:10 (vol/vol) ratio. One microgram of peptide sample was injected to an Eksigent 1D+ Nano LC systems (Eksigent, Dublin) and analyzed in a 5600 TripleTOF mass spectrometer (SCIEX) in SWATH mode as described previously [18]. We reduced the LC gradient to 45min (Supplementary Figure 1), and the SWATH acquisition scheme to 67 variable windows (Supplementary Table 2). The other SWATH parameters were set exactly as in our previous

study [18], except that the ion accumulation time for each SWATH window was 40ms. Ion accumulation time for peptide precursors was set at 50ms. The 38 samples were injected into the mass spec in randomized sequence once and then the same sequence was injected again to obtain a duplicate analysis. After each gradient, the column was washed twice using ramping gradient to minimize carry-over (Supplementary Figure 1). Mass calibration using beta-gal was performed every fourth injection.

OpenSWATH analysis

OpenSWATH (version 2014-12-01-154112) [25] was performed as described previously [18] except that a pan-human library (32 windows version) [23] was used to search the SWATH data in the iPortal platform (Supplementary Table 3, in our laboratory the entire pipeline was streamlined as an iPortal workflow) [26]. The details of the pan-human library, containing 139,449 peptides from 10,316 proteins, have been described previously [23]. Briefly, SWATH wiff files were converted into mzXML files using ProteoWizard msconvert [27]. The mzXML file was then converted to mzML file using OpenMS tool FileConverter [28]. OpenSWATH was performed using the tool OpenSWATHWorkflow with input files including the mzXML file, the TraML pan-human library file, and TraML file for iRT peptides [25]. Trypsin was selected as the proteotypic enzyme. The assays were identified with a false discovery rate below 0.1%. High confidence peptide features from different samples were aligned using the algorithm TTransition of Identification Confidence (TRIC) [29]. Label-free protein-level quantification was performed using the function ProteinInference from the R package aLFQ [30] on output from OpenSWATH.

Pathway and process enrichment analysis by Metascape

Pathway and process enrichment analysis of the differentially abundant proteins were performed using Metascape (<http://metascape.org>). Terms with P value < 0.01, minimum count 3, and enrichment factor > 1.5 (enrichment factor is the ratio between observed count and the count expected by chance) were collected and grouped into clusters based on their membership similarities. To further capture the relationship among terms, a subset of enriched terms was selected and rendered as a network plot, where terms with similarity > 0.3 were connected by edges.

Protein-protein interaction enrichment analysis was carried out for each given gene list (<http://metascape.org>). The resultant network contained the subset of proteins that form physical interactions with at least another list member. Molecular Complex Detection (MCODE)

algorithm [31] was applied to identify densely connected network components. Pathway and biological process enrichment analysis was applied to each MCODE components independently and the three best-scoring (by P value) terms were retained as the functional description of the corresponding components.

IHC staining of HCC tissues

For validation by IHC, three pairs of tissue sections (benign and tumor) from an independent HCC patient cohort which was collected very recently (Union Hospital, Wuhan, China, Supplementary Table 4) were cut out, fixed in formalin and stained using anti-MCM7 antibody (11225-1-AP, Proteintech China, Wuhan). This study was approved by Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology.

Statistics

In volcano plots, paired Student's *t*-test was used to compute P values followed by Benjamini & Hochberg correction in R.

Data deposition

All the SWATH Data files, library and original OpenSWATH results are deposited in PRIDE project [32]: PXD004873. Reviewer account details: Username: reviewer82662@ebi.ac.uk; Password: WTpBL22V

(Note for reviewers: Prior to publication, this data set only be found in PRIDE after the reviewer logs into PRIDE using the username and password provided above. This data set cannot be found without logging in.)

Results and Discussion

Identification of peptides and proteins by PCT-SWATH

To start we extracted proteins from 38 biopsy-level liver tissue samples and produced tryptic peptides using the PCT method [18, 24]. We measured the weight of each wet tissue sample and the total amount of peptide mass generated from the tissue and computed the peptide yield in weight percent per milligram of tissue. The results shown in Figure 1A indicate that the peptide yield for the 38 samples varied from 1.0% to 9.5%, with an average value of 6.0%. 59.6 µg of total peptide mass per milligram wet liver tissue were obtained on average. The peptide yield in tumor tissue (4.9% on average) was lower than that of benign tissue (8.1% on average, P value = 0.00037) (Figure 1A). The results further showed that tumor samples

exhibited more variability in terms of peptide yield compared to the benign tissue samples. The CV of peptide yield was 15.8% for benign tissue samples ($n = 19$), 40.7% for tumor samples ($n = 19$), and 32.9% overall ($n = 38$) (Figure 1B), demonstrating the heterogeneity of liver tumors.

Each peptide sample was then analyzed using SWATH-MS in technical duplicate, using a 45-min LC gradient and a 67 variable SWATH window acquisition scheme (Supplementary Table 2) which was optimized based on the ion density map of human proteins extracted from HCC tissues. With this SWATH window setting, up to 20 samples were analyzed per MS instrument per 24 hours. Mass calibration using beta-gal was performed every fourth injection. The MS analysis of the 76 SWATH runs were completed in about 4 working days.

After searching the resulting raw data files against the pan-human SWATH assay library [23], 2,579 unique proteotypic SwissProt proteins with 11,787 unique peptides were identified at a false discovery rate (FDR) below 1% using OpenSWATH [23] (Supplementary Table 5). Comparable numbers of proteins and peptides were quantified from both benign and tumor tissue samples (Figure 2A and Figure 2B). In this study, the PCT-SWATH throughput was further increased compared to the previous work [18]. The protein numbers identified per sample were slightly higher than those obtained from human kidney tissues [18].

Variability of protein expression between tumor tissue and its adjacent benign pair

To evaluate the quantitative accuracy of the method, we computed the global Pearson correlation of the two technical replicates at the peptide level for all samples after removing missing values. The R^2 was 0.9729 (Figure 2C), indicating that our label-free method could quantify peptides with high degree of reproducibility. The global view of protein quantification is shown in Figure 2D. It showed that all the technical replicates were clustered together respectively.

We then analyzed the overall variation of protein abundance for all 2,579 quantified proteins from tumor tissues, adjacent benign tissues and all samples, respectively (Figure 3A). The data showed that the CV of tumor tissue samples was slightly greater than that of adjacent benign tissues, indicating more overall variability of protein expression in tumors. We next checked the relationship between protein abundance and the overall variation among these samples and observed that low abundance proteins exhibited a higher degree of overall variability (Figure 3B, 3C, 3D). In Figure 2C, proteins of low abundance demonstrated comparable reproducibility with high abundance proteins, therefore we mainly attributed the higher variability in tumor tissue to biological causes.

Differentially regulated proteins characterized in the HCC cohort

We further explored differentially expressed proteins between tumor and benign samples. A total of 541 proteins showed significant differential abundance (Figure 4A). Of these, 381 proteins showed increased abundance and 160 proteins showed decreased abundance in tumor tissues compared to benign tissues. Pathway enrichment analysis of the regulated proteins in HCC tumor tissues was performed by Metascape. In total, 24 pathways were distinguished to be activated and 13 pathways were found to be suppressed, as shown in Supplementary Table 6. We found that proteins with increased abundance were highly enriched in mRNA related processes including mRNA metabolic process, peptide metabolic process, ribonucleoprotein complex biogenesis, and nuclear transport. The proteins with decreased abundance were found to be mostly involved in small molecule metabolic processes such as small molecule catabolic process, monocarboxylic acid metabolic process, and carbon metabolism. Next, we carried out protein-protein interaction network analysis using Metascape. The resultant network contains the subset of proteins that form physical interactions with at least another protein showing altered abundance.

Heat shock proteins (HSP) were found to be upregulated in our dataset, such as HSPA1A, HSPA1L, HSPA4, HSPA8, HSPB1, HSP90AB1 (Figure 4B, Supplementary Table 6). HSP proteins constitute a family of highly conserved stress response proteins that can protect cells and induce them to repair damage caused by a variety of stress [33]. Several members of the HSP family have been reported to be overexpressed in a wide range of human cancers [34]. It has been demonstrated that 14-3-3 σ protein induces HSP70 via a β -catenin/HSF-1-dependent pathway, which consequentially modulates HCC. The HSF-1/Hsp70-1 protein complex might be of prognostic value for HCC [35]. We didn't detect HSP70 in this dataset, but a series of heat shock proteins have been characterized to be over expressed in this cohort.

We also observed that proteins related to mitochondrial translation, specifically the mitochondrial ribosomal proteins (MRPs), were among the most strongly upregulated protein clusters (Figure 4C). Tumor initiation and progression in cancer cells involve the development of mechanisms to inhibit apoptosis at multiple stages, whereas mitochondria play crucial roles in the induction of apoptosis and programmed cell death. This paradigm is central to malignant cellular transformation because altered mitochondrial function and defective apoptosis are well-known hallmarks of cancer cells. It has also been reported that the expression of genes encoding MRPs, mitoribosome assembly factors and mitochondrial translation factors is

modulated in multiple cancers, which is linked to tumorigenesis and metastasis [36]. Mini-chromosome maintenance complex component 7 (MCM7) was also found to be overexpressed (Figure 4C). MCM7 plays an essential role in initiating DNA replication. DNA replication is a central process in cell proliferation, while aberrant DNA replication is indicated to be a driving force of oncogenesis. It has been reported that overexpressed MCM7 is associated with a poor prognosis of HCC patients, and MCM7 promotes HCC cell proliferation via upregulating MAPK-cyclin D1 pathway both *in vitro* and *in vivo* [37]. In addition, the hallmark MYC targets proteins, and the cell cycle checkpoint proteins involved in G2/M transitions, are also among the upregulated protein networks.

Small molecule catabolic processes characterized the group of proteins that showed the strongest reduction of abundance in the HCC samples tested. Our data showed that proteins involved in glycogen metabolism pathway were suppressed in HCC samples (Figure 4D). These proteins include the glycogen synthesis enzymes including phosphoglucomutase-1 (PGM1), and UTP-glucose-1-phosphate uridylyltransferase (UGP2), as well as the glycogen degradation enzyme liver glycogen phosphorylase (PYGL). They are downregulated in tumor tissues compared to their matched benign controls, indicating the reprogramming of glycogen metabolism in HCC which suppresses the turn-over of glycogen. Glycogen metabolism is a key pathway induced by hypoxia, and is necessary for optimal glucose utilization. In mammalian cells, glycogen is the main storage form of glucose. The liver accumulates glycogen in large amounts and releases it slowly to maintain blood glucose levels [38]. Recently glycogen metabolism has become a recognized feature of cancer cells. As to HCC, it has been reported that inhibition of glycogen synthase kinase-3 β (GSK-3 β) suppressed the growth and induced caspase-dependent apoptosis in HCC cells, and GSK-3 β may become a promising therapeutic target for HCC [39-41]. In our dataset, glycogen metabolism pathway was found to be suppressed in HCC patients. Besides, our data did not indicate whether this reprogramming changed the direction of glycogen storage.

The ethanol metabolism was also found to be downregulated. Aldehyde dehydrogenase (ALDH) is a gene superfamily that is responsible for the detoxification of biogenic and xenogenic aldehydes. Both ALDH2 and ALDH1B1 proteins were significantly downregulated in our dataset (Figure 4D). Similarly, a previous study by Park *et al* revealed increased detection of ALDH-3 but decreased detection of ALDH-2 when measured by 2D electrophoresis and mass spectrometry in a group of ten cases with HCC [42]. ALDH1 and ALDH2 are the most important enzymes for aldehyde oxidation. The ethanol detoxifying pathway in humans occurs mainly in

the liver and is carried out by two enzymatic steps. In the first step ethanol is metabolized quickly by alcohol dehydrogenase (ADH) to generate acetaldehyde, and the latter is then metabolized by the mitochondrial ALDH2 to acetate [43]. ALDH2 is expressed ubiquitously in all tissues but is most abundant in the liver and found in high amounts in organs that require high mitochondrial oxidative phosphorylation. The downregulation of ALDH2 in liver tumor tissue will likely accumulate toxin acetaldehyde and promote tumorigenesis.

In summary, liver is the pivotal organ for molecular synthesis and the metabolic hub of human beings. Our data showed that the molecular synthesis related pathways were enhanced, while most major metabolic processes were suppressed in the HCC tissues tested, suggesting a biochemical imbalance in HCC cells.

Proteins with altered abundance in HCC patients with high serum AFP

Next, we explored the molecular pathogenesis of HCC by associating protein expression with various clinic-pathological characteristics of HCC. AFP is the most widely used tumor biomarker currently available for early detection of HCC, and the widely accepted threshold of serum AFP is 20ng/ml. It has been reported that at this threshold, serum AFP had a sensitivity of 41-65% and specificity of 80-94% [44]. We grouped the 19 HCC patients in our cohort into two groups, according to an AFP-cutoff of 20ng/mL (Supplementary Table 1), computed the fold change of protein expression between tumor and benign tissues, and checked the protein regulation pattern of 11 patients with high AFP level (> 20ng/ml) and 6 patients with low AFP level (< 20 ng/ml), respectively. Two patients without serum AFP examination value at surgical operation were excluded from these analyses. Although a considerable proportion of HCC patients in the cohort do not have elevated serum AFP, and serum AFP can increase in patients with diseases other than HCC [4], high expression of AFP in serum correlated with high cell proliferation, high angiogenesis and low apoptosis and was associated with poor prognosis of HCC [45]. Indeed, we found in our data set that the low serum AFP level was closely related to higher survival rate of the patients (5 out of 6 patients have recurrence free survival), especially when there was no portal vein thrombosis detected by CT at diagnosis, as is shown in Supplementary Table 1. On the other hand, in the AFP-high patient group, 6 of 11 patients were dead and 2 patients suffered from tumor recurrence, displaying much poorer survival (Supplementary Table 1). Nevertheless, statistical analysis was not feasible due to small sample size.

With respect to the 11 AFP-high patient group, 419 proteins showed increased abundance and 192 proteins showed decreased abundance (P value < 0.05) in tumor tissues compared to tumor tissues (Figure 4E). In the AFP-low subgroup of 6 patients, our data showed that no protein significantly regulated between the tumor and benign tissues, probably due to the small sample size (Figure 4F). Pathway enrichment and network analysis of the differentially expressed proteins from the AFP-high subgroup revealed similar regulation pattern with the 19 HCC patients group.

In the AFP-high subgroup, we identified additional downregulated proteins including methylene tetrahydrofolate dehydrogenase 1 (MTHFD1) (Figure 4G), which is an enzyme identified in the one-carbon (1C) metabolism pathway. Another enzyme in the pathway, *i.e.* serine hydroxymethyltransferase (SHMT1), was also downregulated (Figure 4D). The data suggest reprogramming of one-carbon metabolism in HCC. This is consistent with the Liverome study reporting that downregulation of MTHFD1 is prevalent in proliferative and poorly differentiated HCCs [46].

We also observed downregulation of a tumor suppressor protein regucalcin (RGN) (Figure 4H) which modulates multiple protein kinases and phosphatases in cancers including HCC [47]. Transcripts analysis has reported that prolonged survival in HCC patients is associated with increased RGN gene expression [48]. Our data at protein level further consolidate the potential roles of RGN in HCC.

Two proteins involved in lipid metabolism showed decreased expression in the AFP-high HCC tissues. They are ACAA2 and ANXA6 (Figure 4H). ACAA2 is a mitochondrial enzyme involved in the fatty acid metabolism pathway. ANXA6 is a lipid-binding protein highly expressed in the liver, regulating cholesterol homeostasis and signaling pathways with a role in liver physiology. A recent study found that ANXA6 was downregulated in HCC tissues compared to benign tissues in 18 HCC patients [49]. The data indicate dysfunction of lipid metabolism in HCC tissues.

Quantitative proteomics analysis of Hepatitis B virus positive (HBV+) HCC subgroup

Chronic HBV is the most common cause of HCC in China. Approximately 80% of HCC develops from liver cirrhosis, which predominantly progresses from HBV in China [50]. We found that 106 and 75 proteins showed increased or decreased abundance, respectively, in 11 HCC patients with HBV infection (Figure 4I, Supplementary Table 6) compared to those without HBV infection. Subsequent pathway and protein interaction analysis by Metascape revealed a

number of interesting protein clusters that showed a close relationship to liver cirrhosis in patients with chronic HBV infection. Pyruvate kinase M (PKM) was found to be upregulated in this HBV⁺ subgroup (Figure 4J). PKM is a protein kinase that catalyzes the final step in glycolysis by transferring the phosphate from phosphoenolpyruvate (PEP) to ADP, thereby generating pyruvate and ATP [51]. PKM has two isoforms, PKM1 and PKM2. PKM2 plays an important role in metabolic alterations related to inflammation and cancer. The ATP generation by PKM2, unlike mitochondrial respiration, is independent of oxygen and thus allows tumor cells to grow in hypoxic conditions. Cancer cells are characterized by high glycolytic rates to support energy regeneration and anabolic metabolism, along with the high expression of PKM2 [52]. It has been shown in a recent report that the anti-apoptotic protein poly (ADP-ribose) polymerase (PARP)14 promotes aerobic glycolysis in HCC by maintaining low activity of PKM2, and that the PARP14-JNK1-PKM2 regulatory axis links apoptosis and metabolism [53]. In agreement with our finding that HBV⁺ tumors expressed higher level of PKM, there is another report showing that high PKM2 expression was more frequently found in cirrhotic liver caused by HBV infection than in non-cirrhotic liver, and that PKM2 overexpression was associated with poor survival rates in HCC patients with cirrhotic liver (CL) [54].

Cytidine triphosphate synthase 1 (CTPS1) is another example of a protein that showed higher expression in the HBV⁺ HCC subgroup in this study compared to patients without HBV infection (Figure 4J). CTPS1 catalyzes the rate-limiting step in the *de novo* CTP synthetic pathway, in which a UTP is converted into CTP with the consumption of glutamine and ATP. In fact, upregulated CTPS1 activity has been observed in multiple human and rodent tumors [55], promoting tumor transformation and progression [56]. Recently, researchers focus on the CTPS cytoophidium that is a filamentous intracellular macrostructure aggregated by CTPS, and is able to promote the activity of CTPS [57]. A recent study examined the presence of CTPS cytoophidia in tumor and the adjacent benign tissues of HCC patients, and found that many cytoophidia were observed in 28% of the HCC tumor samples but not the adjacent hepatocyte population [58]. In addition, they found that the high expression of HSP90 was correlated with the presence of CTPS cytoophidia, which is consistent to our findings that heat shock proteins (HSPA1L and HSPA8) and CTPS1 are all upregulated (Figure 4J).

On the other hand, both cystathionine β -synthase (CBS) and cystathionine- γ -lyase (CTH) were found at decreased abundance in the HCC samples tested (Figure 4K). CBS and CTH are abundant proteins in the liver, endogenously produce hydrogen sulfide (H₂S), a gasotransmitter modulating synaptic transmission, vasorelaxation, angiogenesis, inflammation, and cellular

bioenergetics. Regulation of H₂S influences lipid metabolism, glucose metabolism, oxidative stress, and mitochondrial bioenergetics [59]. Suppression of CBS transcripts has been reported in a study of 120 HCC patients compared to non-cancerous liver tissue [60]. Reduced CBS transcripts expression is significantly correlated with high tumor stage, high Edmondson grade, and high AFP level [60]. However, neither the CBS nor the CTH protein level has yet been reported to be investigated in HCC tissue. Our data suggest downregulation of H₂S production via suppressing CBS and CTH in HCC tissues.

Fructose-bisphosphate aldolase B (ALDOB) is an enzyme for glucose and fructose metabolism. In a cohort of 313 HCC patients, ALDOB was found downregulated in HCC tissue using tissue microarray and IHC technologies, and its downregulation was significantly correlated with HCC progression [61]. Furthermore, ALDOB has been found to be a binding protein of the S region of the HbsAg [62]. Our data showed that ALDOB expression was suppressed in HBV⁺ HCC patients (Figure 4L), suggesting a potentially interesting mechanistic linkage between HCC and HBV infection, which might be utilized as therapeutic target.

Summary of regulated proteins in HCC samples

We summarized all differentially expressed proteins in HCC samples including the AFP-high group and HBV⁺ group (Figure 5A). Interestingly, the proteins with increased expression concentrated in three types of pathways: a) production of DNA, mRNA and protein; b) oncogenic signaling pathways and apoptosis; c) immune response. This suggests enhanced proliferation of tumor cells in the malignant tissues. In contrast, proteins participated in various metabolism pathways were down-regulated, indicating the suppression of metabolic functionality in liver during oncogenesis.

From literature review of all these regulated proteins in Figure 5A, we identified 16 proteins of reported clinical relevance to HCC (Figure 5B). We calculated the expression level of these 16 proteins between tumor and the adjacent benign tissue, as was shown in Figure 5B, the log₁₀ MS intensity, with the P value for each pair of comparison between benign and tumor.

Among these potential candidates, MTHFD1, RGN, CBS, and CTH were previously reported to be dysregulated in HCC as transcripts at gene level, as discussed above. The protein expression level of MCM7, ALDH2, ANXA6, PKM2, CTPS and ALDOB were reported to be changed in HCC in other studies. The others haven't been reported yet.

Detection of MCM7 in HCC samples using IHC

To verify whether findings from our SWATH data are consistent with IHC reports, we further applied MCM7 IHC staining to three additional patients who accepted surgical operation very recently (Supplementary Table 4) and observed the up-regulation of MCM7 in liver tumors (Figure 6). The data collectively suggest that upregulation of MCM7 might play crucial roles in liver cancer progression.

Conclusion

In conclusion, here we report an optimized PCT-SWATH workflow enabling analysis of clinical tissue specimens with increased sample throughput without compromise of quantitative accuracy and proteomic coverage. Our study identified a few regulated proteins in this HCC cohort. Proteins with increased abundance are mainly related to mass production, oncogenic signaling and immunity, whereas metabolic proteins are shown with lower expression. We identified 16 dysregulated proteins of reported clinical relevance to HCC. Upregulation of MCM7 in tumor tissue samples was observed using IHC in additional patients. The study demonstrates that the PCT-SWATH methodology has potential to be practically applied in clinical research to analyze tissue samples in high-throughput.

Author contributions

T.G. conceived and coordinated the project. P.S. procured the tissue samples and helped with finding the relevant matching tumor/benign sample pairs. J.Z., and C.L. punched all tissue blocks. J.Z., C.L., W.X. and Y.Z. performed the PCT-based sample preparation. Y.Z and T.G. performed the SWATH-MS data acquisition. P.S., J.Z. and C.L. provided clinical follow-up data. X.D. supervised Y. S., X. Y. to evaluate the MCM7 IHC staining pictures. J.Z., Y.Z., Q.Z., L.Y. T.Z., T.G. performed the data analysis. Y.Z., J.Z., R.A., S.H., T.G. wrote the manuscript with inputs from all co-authors. R.A., S.H. and T.G. supervised the project.

Acknowledgements

This study was supported the National Natural Science Foundation, P.R China (NNSF/81200348/2013), the Swiss National Science Foundation (SNF 166435 MitoModules to R.A) ERC PROTEOMICS4D (project-no 670821). Y. Z. was supported by a fellowship from SignalX (SystemsX RTD 2013/156) and Eu PRECISE (project-no 668858).

Conflict of Interest

R.A. holds shares of Biognosys AG, which operates in the field covered by the article. The research groups of R.A. and T.G. are supported by AB SCIEX, which provides access to prototype instrumentation, and Pressure Biosciences, which provides access to advanced sample preparation instrumentation.

References

1. El-Serag, H.B. and K.L. Rudolph, *Hepatocellular carcinoma: epidemiology and molecular carcinogenesis*. Gastroenterology, 2007. **132**(7): p. 2557-76.
2. Yau, T., et al., *Development of Hong Kong Liver Cancer staging system with treatment stratification for patients with hepatocellular carcinoma*. Gastroenterology, 2014. **146**(7): p. 1691-700 e3.
3. Ferrin, G., et al., *Biomarkers for hepatocellular carcinoma: diagnostic and therapeutic utility*. Hepat Med, 2015. **7**: p. 1-10.
4. Schutte, K., et al., *Current biomarkers for hepatocellular carcinoma: Surveillance, diagnosis and prediction of prognosis*. World J Hepatol, 2015. **7**(2): p. 139-49.
5. Bruix, J., M. Sherman, and D. American Association for the Study of Liver, *Management of hepatocellular carcinoma: an update*. Hepatology, 2011. **53**(3): p. 1020-2.
6. European Association For The Study Of The, L., R. European Organisation For, and C. Treatment Of, *EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma*. J Hepatol, 2012. **56**(4): p. 908-43.
7. Omata, M., et al., *Asian Pacific Association for the Study of the Liver consensus recommendations on hepatocellular carcinoma*. Hepatol Int, 2010. **4**(2): p. 439-74.
8. Saffroy, R., et al., *New perspectives and strategy research biomarkers for hepatocellular carcinoma*. Clin Chem Lab Med, 2007. **45**(9): p. 1169-79.
9. Nie, W., et al., *Advanced mass spectrometry-based multi-omics technologies for exploring the pathogenesis of hepatocellular carcinoma*. Mass Spectrom Rev, 2016. **35**(3): p. 331-49.
10. Kimhofer, T., et al., *Proteomic and metabonomic biomarkers for hepatocellular carcinoma: a comprehensive review*. Br J Cancer, 2015. **112**(7): p. 1141-56.
11. Abdalla, M.A. and Y. Haj-Ahmad, *Promising Urinary Protein Biomarkers for the Early Detection of Hepatocellular Carcinoma among High-Risk Hepatitis C Virus Egyptian Patients*. J Cancer, 2012. **3**: p. 390-403.
12. Zhang, J., et al., *In-depth proteomic analysis of tissue interstitial fluid for hepatocellular carcinoma serum biomarker discovery*. Br J Cancer, 2017. **117**(11): p. 1676-1684.
13. Le Faouder, J., et al., *Imaging mass spectrometry provides fingerprints for distinguishing hepatocellular carcinoma from cirrhosis*. J Proteome Res, 2011. **10**(8): p. 3755-65.
14. Marquardt, C., et al., *MALDI imaging-based classification of hepatocellular carcinoma and non-malignant lesions in fibrotic liver tissue*. Z Gastroenterol, 2015. **53**(1): p. 33-9.
15. Li, C., et al., *Quantitative proteomics reveal up-regulated protein expression of the SET complex associated with hepatocellular carcinoma*. J Proteome Res, 2012. **11**(2): p. 871-85.

16. Megger, D.A., et al., *Proteome Analyses of Hepatocellular Carcinoma*. J Clin Transl Hepatol, 2014. **2**(1): p. 23-30.
17. Naboulsi, W., et al., *Quantitative Tissue Proteomics Analysis Reveals Versican as Potential Biomarker for Early-Stage Hepatocellular Carcinoma*. J Proteome Res, 2016. **15**(1): p. 38-47.
18. Guo, T., et al., *Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps*. Nat Med, 2015. **21**(4): p. 407-13.
19. Shao, S., et al., *Reproducible Tissue Homogenization and Protein Extraction for Quantitative Proteomics Using MicroPestle-Assisted Pressure-Cycling Technology*. J Proteome Res, 2016. **15**(6): p. 1821-9.
20. Shao, S., et al., *Minimal sample requirement for highly multiplexed protein quantification in cell lines and tissues by PCT-SWATH mass spectrometry*. Proteomics, 2015. **15**(21): p. 3711-21.
21. Gillet, L.C., et al., *Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis*. Mol Cell Proteomics, 2012. **11**(6): p. O111 016717.
22. Guo, T., et al., *Multi-region proteome analysis quantifies spatial heterogeneity of prostate tissue biomarkers*. Life Sci Alliance, 2018. **1**(2).
23. Rosenberger, G., et al., *A repository of assays to quantify 10,000 human proteins by SWATH-MS*. Sci Data, 2014. **1**: p. 140031.
24. Zhu, Y. and T. Guo, *High-Throughput Proteomic Analysis of Fresh-Frozen Biopsy Tissue Samples Using Pressure Cycling Technology Coupled with SWATH Mass Spectrometry*. Methods Mol Biol, 2017.
25. Rost, H.L., et al., *OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data*. Nat Biotechnol, 2014. **32**(3): p. 219-23.
26. Peter Kunszt, et al., *iPortal: the swiss grid proteomics portal: Requirements and new features based on experience and usability considerations*. 2014: p. 10.1002/cpe.3294.
27. Adusumilli, R. and P. Mallick, *Data Conversion with ProteoWizard msConvert*. Methods Mol Biol, 2017. **1550**: p. 339-368.
28. Sturm, M., et al., *OpenMS - an open-source software framework for mass spectrometry*. BMC Bioinformatics, 2008. **9**: p. 163.
29. Rost, H.L., et al., *TRIC: an automated alignment strategy for reproducible protein quantification in targeted proteomics*. Nat Methods, 2016. **13**(9): p. 777-83.
30. Rosenberger, G., et al., *aLFQ: an R-package for estimating absolute protein quantities from label-free LC-MS/MS proteomics data*. Bioinformatics, 2014. **30**(17): p. 2511-3.
31. Bader, G.D. and C.W. Hogue, *An automated method for finding molecular complexes in large protein interaction networks*. BMC Bioinformatics, 2003. **4**: p. 2.
32. Vizcaino, J.A., et al., *2016 update of the PRIDE database and its related tools*. Nucleic Acids Res, 2016. **44**(D1): p. D447-56.
33. Zhao, Y.J., Q. Ju, and G.C. Li, *Tumor markers for hepatocellular carcinoma*. Mol Clin Oncol, 2013. **1**(4): p. 593-598.
34. Ciocca, D.R. and S.K. Calderwood, *Heat shock proteins in cancer: diagnostic, prognostic, predictive, and treatment implications*. Cell Stress Chaperones, 2005. **10**(2): p. 86-103.

35. Fang, F., R. Chang, and L. Yang, *Heat shock factor 1 promotes invasion and metastasis of hepatocellular carcinoma in vitro and in vivo*. *Cancer*, 2012. **118**(7): p. 1782-94.
36. Kim, H.J., P. Maiti, and A. Barrientos, *Mitochondrial ribosomes in cancer*. *Semin Cancer Biol*, 2017. **47**: p. 67-81.
37. Qu, K., et al., *MCM7 promotes cancer progression through cyclin D1-dependent signaling and serves as a prognostic marker for patients with hepatocellular carcinoma*. *Cell Death Dis*, 2017. **8**(2): p. e2603.
38. Zois, C.E. and A.L. Harris, *Glycogen metabolism has a key role in the cancer microenvironment and provides new targets for cancer therapy*. *J Mol Med (Berl)*, 2016. **94**(2): p. 137-54.
39. Cervello, M., et al., *Pivotal roles of glycogen synthase-3 in hepatocellular carcinoma*. *Adv Biol Regul*, 2017. **65**: p. 59-76.
40. Desbois-Mouthon, C., et al., *Dysregulation of glycogen synthase kinase-3beta signaling in hepatocellular carcinoma cells*. *Hepatology*, 2002. **36**(6): p. 1528-36.
41. Zhang, N., et al., *Glycogen synthase kinase-3beta inhibition promotes lysosome-dependent degradation of c-FLIPL in hepatocellular carcinoma*. *Cell Death Dis*, 2018. **9**(2): p. 230.
42. Park, K.S., et al., *Proteomic alterations of the variants of human aldehyde dehydrogenase isozymes correlate with hepatocellular carcinoma*. *Int J Cancer*, 2002. **97**(2): p. 261-5.
43. Chen, C.H., et al., *Targeting aldehyde dehydrogenase 2: new therapeutic opportunities*. *Physiol Rev*, 2014. **94**(1): p. 1-34.
44. Lou, J., et al., *Biomarkers for Hepatocellular Carcinoma*. *Biomark Cancer*, 2017. **9**: p. 1-9.
45. Yokoo, H., et al., *Proteomic signature corresponding to alpha fetoprotein expression in liver cancer cells*. *Hepatology*, 2004. **40**(3): p. 609-17.
46. Lee, L., et al., *Liverome: a curated database of liver cancer-related gene signatures with self-contained context information*. *BMC Genomics*, 2011. **12 Suppl 3**: p. S3.
47. Yamaguchi, M., *Involvement of regucalcin as a suppressor protein in human carcinogenesis: insight into the gene therapy*. *J Cancer Res Clin Oncol*, 2015. **141**(8): p. 1333-41.
48. Yamaguchi, M., et al., *Prolonged survival in hepatocarcinoma patients with increased regucalcin gene expression: HepG2 cell proliferation is suppressed by overexpression of regucalcin in vitro*. *Int J Oncol*, 2016. **49**(4): p. 1686-94.
49. Meier, E.M., et al., *Annexin A6 protein is downregulated in human hepatocellular carcinoma*. *Mol Cell Biochem*, 2016. **418**(1-2): p. 81-90.
50. Zhang, Y., et al., *Epidemiology of hepatitis B and associated liver diseases in china*. *Chin Med Sci J*, 2013. **27**(4): p. 243-8.
51. Dong, G., et al., *PKM2 and cancer: The function of PKM2 beyond glycolysis*. *Oncol Lett*, 2016. **11**(3): p. 1980-1986.
52. Iqbal, M.A., et al., *Pyruvate kinase M2 and cancer: an updated assessment*. *FEBS Lett*, 2014. **588**(16): p. 2685-92.
53. Iansante, V., et al., *PARP14 promotes the Warburg effect in hepatocellular carcinoma by inhibiting JNK1-dependent PKM2 phosphorylation and activation*. *Nat Commun*, 2015. **6**: p. 7882.

54. Liu, Y., et al., *Clinicopathological and prognostic significance of PKM2 protein expression in cirrhotic hepatocellular carcinoma and non-cirrhotic hepatocellular carcinoma*. Sci Rep, 2017. **7**(1): p. 15294.
55. Kizaki, H., et al., *Increased cytidine 5'-triphosphate synthetase activity in rat and human tumors*. Cancer Res, 1980. **40**(11): p. 3921-7.
56. Williams, J.C., et al., *Increased CTP synthetase activity in cancer cells*. Nature, 1978. **271**(5640): p. 71-3.
57. Liu, J.L., *The Cytoophidium and Its Kind: Filamentation and Compartmentation of Metabolic Enzymes*. Annu Rev Cell Dev Biol, 2016. **32**: p. 349-372.
58. Chang, C.C., et al., *CTP synthase forms the cytoophidium in human hepatocellular carcinoma*. Exp Cell Res, 2017. **361**(2): p. 292-299.
59. Mani, S., et al., *Hydrogen sulfide and the liver*. Nitric Oxide, 2014. **41**: p. 62-71.
60. Kim, J., et al., *Expression of cystathionine beta-synthase is downregulated in hepatocellular carcinoma and associated with poor prognosis*. Oncol Rep, 2009. **21**(6): p. 1449-54.
61. Tao, Q.F., et al., *Aldolase B inhibits metastasis through Ten-Eleven Translocation 1 and serves as a prognostic biomarker in hepatocellular carcinoma*. Mol Cancer, 2015. **14**: p. 170.
62. Wu, J., et al., *ALDOB acts as a novel HBsAg-binding protein and its coexistence inhibits cisplatin-induced HepG2 cell apoptosis*. Crit Rev Eukaryot Gene Expr, 2014. **24**(3): p. 181-91.

Figure legend

Figure 1. Performance of PCT-assisted peptide preparation. (A) Peptide yield in two types of liver tissue samples. Y-axis shows the peptide yield in percentage of peptide (μg) per milligram (mg) tissue. (B) CV of peptide yield in two types of tissues and across all samples.

Figure 2. Peptide and protein identification and quantification. (A) Number of peptides identified in each sample (76 SWATH runs of 38 peptide samples from 19 pairs of HCC patient tissues). **(B)** Number of proteins identified in each sample. **(C)** Global Pearson correlation of technical replicates of peptide quantification for samples in all samples. Values are log₂ transformed protein intensity. Missing values were excluded from this analysis. **(D)** Heatmap overview of protein abundance patterns in 76 samples. Protein intensity values were log₁₀ transformed and subjected to unsupervised-clustering in two axes. (P, patient ID; N, benign tissue; T, tumor tissue; a, the first technical replicate; b, the second technical replicate; for example, P1Na, the technical replicate 'a' of benign tissue from patient No. 1.)

Figure 3. Variability of protein expression in liver tissues. (A) Distribution of protein abundance variability, quantified as percentage of CV values, in all kinds of tissue samples. **B, C** and **D** show correlation of CV of protein abundance variability and protein abundance of all samples, benign samples and tumor samples, respectively.

Figure 4. Significantly regulated proteins in tumor versus benign tissues from the 19 HCC patient cohort (ALL), the high blood AFP level subgroup (AFP_High) of 11 HCC patients, the low blood AFP level subgroup (AFP_Low) of 6 HCC patients, and the HBV infected subgroup (HBV⁺) of 11 HCC patients, respectively. (A) Volcano plot of proteins expressed in tumor tissue samples and their matched benign samples from the 'ALL' group. Each data point represents the ratio of tumor to adjacent benign tissue observed for a protein. The horizontal line at $-\log_{10}(P) = 1.3$ indicates $P = 0.05$. Data points above the line ($P < 0.05$) are considered to be significantly regulated. The two vertical lines at $\log_2(T/N) = \pm 1$ where the fold change is equivalent to 2 indicate the threshold for up-/down- regulation. The upper left quadrant ($P < 0.05$ and fold change < 0.5) represent the downregulated proteins whereas the upper right quadrant indicates the upregulated proteins in tumor tissue. N, benign; T, tumor. **(B)** and **(C)** enriched significantly upregulated protein networks for the 19 HCC patient cohort. **(D)** enriched significantly downregulated protein network for the 19 HCC patient cohort. **(E)** Volcano plots of proteins expressed in tumor tissue samples and their matched benign samples in the 11 AFP-high patients group. **(F)** Volcano plots of proteins expressed in tumor tissue samples and their matched benign samples in the 6 AFP-low patients group. **(G)** and **(H)** show the enriched significantly downregulated protein network in the AFP-high patient group. **(I)** Volcano plots of proteins expressed in tumor tissue samples and their matched benign samples in 11 HBV⁺ HCC patients. **(J)** The enriched significantly upregulated protein networks in the HBV positive patient group. **(K)** and **(L)** show the downregulated protein networks in the HBV positive patient group.

Figure 5. Summary of regulated proteins and pathways in HCC tumor tissues compared to matched benign samples. (A) All proteins found regulated in HCC tissues, including AFP-high group and HBV⁺ group, were inspected in literature. Regulated pathways and their component proteins are displayed. **(B)** 16 selected proteins significantly regulated in HCC tumor tissues compared to matched benign samples. For each protein, its expression, after $\log_{10}$ transformation, in benign (N) and tumor (T) tissues are shown in boxplots. Student's *t*-test was used to compute the significance between N and T samples.

A

Up-regulated proteins and pathways

Down-regulated proteins and pathways

DNA replication MCM7 SSRP1 DNA repair RUVBL2 NONO SMC1A SSRP1 Transcription initiation and elongation ROLRC SSRP1 POLR1C PSMC1 Transcription regulation RUVBL2 BTF3 CTNNB1 HSPA8 NONO NCL DDX3X GCN1 mRNA maturation PABPC1 NONO DDX3X HSPA8 Regulation of mRNA stability HSPA1A HSPA8 HSPB1 PABPC1 PSMC1 Translation initiation EIF3F/EIF3H EIF4H PABPC1 RPL8 RPS7 NARS	Translation regulation EIF3F/EIF3H EIF4H HSPB1 MTIF2 DDX3X Posttranslational modification PSMC1 AHSA1 Protein processing in ER HSP90AB1 HSPA1A HSPA1L HSPA8 Cell proliferation SMC1A RCC1 CTNNB1 PAK2 EIF3F/EIF3H Mitochondria transcription transcription MRPL12 Mitochondria Translation MRPL11 MRPL12 MRPL19 MRPL37 MRPL41 MRPL49 MRPL53 MRPS23 MTIF2	Cell-cell adhesion&Cell-cell adherens junction OLA1 AHSA1 RAB1A ABCF3 DDX3X CTNNB1 NC1 PAK2 GCN1 EIF4H HSP90AB1 HSPA1A HSPA8 PKM Stress response GCN1 AHSA1 HSP90AB1 HSPA1L HSPA1A HSPA4 HSPA8 HSPB1 DDX3X MAPK signaling pathway HSPA1A/HSPA1L HSPA8 HSPB1 PAK2 Wnt signaling pathway DDX3X CTNNB1 MARK1 Tumor necrosis factor-mediated signaling pathway HSPB1 HSPA1A PSMC1	Oncogene family RAB1A Pathways in cancer CTNNB1 HSP90AB1 PKM ATPase activity ABCF3 DDX3X OLA1 RUVBL2 HSPA1A HSPA8 PSMC1 Apoptosis PSMC1 EIF3F EIF3H DDX3X CTNNB1 PAK2 RUVBL2 Antigen processing & presentation HSP90AB1 HSPA1A HSPA1L HSPA4 HSPA8 HSPB1 Innate Immune response NONO DDX3X CTNNB1	Amino-acid biosynthesis CTH CBS MTHFD1 OTC ALDOB SHMT1 Pyruvate metabolism ACAT1 ACSS2 ALDH1B1 ALDH2 Carbon metabolism ACAT1 ACSS2 ALDOB RGN SHMT1 Terpenoid backbone biosynthesis ACAT1 FDPS ALDH1B1 Glycolysis/ Gluconeogenesis ACSS2 ALDH2 PGM1 Sucrose metabolism UGP2 PGMM1 PYGL	Antibiotics biosynthesis UGP2 ACAT1 ACAA2 ACSS2 ALDH1B1 ALDH2 ALDOB CTH CBS FDPS OTC PGM1 RGN SHMT1 Fatty acid metabolism ACAA2 ACAT1 Lipid biosynthesis FDPS Fatty acid degradation ACAT1 ACAA2 ALDH1B1 ALDH2 Ethanol oxidation ACSS2 ALDH1B1 ALDH2 Pentose phosphate pathway ALDOB PGM1 RGN Glycogen biosynthetic process UGP2 PGM1
---	---	---	---	--	--

B

Figure 6. Anti-MCM7 IHC staining (20X and 40X) of benign and tumor tissues from three HCC patients in pairwise. Note: P2-1 refers to the 1st patient from the 2nd cohort as shown in Supplementary Table 4. The left column shows the staining for benign samples while the right column shows the staining for tumor samples.